

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁵ : C12N 15/56, 9/42		A2	(11) International Publication Number: WO 94/21801
			(43) International Publication Date: 29 September 1994 (29.09.94)
(21) International Application Number: PCT/US94/02657			(81) Designated States: CA, FI, JP, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).
(22) International Filing Date: 11 March 1994 (11.03.94)			
(30) Priority Data: 08/032,848 17 March 1993 (17.03.93) US			
(71) Applicant: GENENCOR INTERNATIONAL, INC. [US/US]; 180 Kimball Way, South San Francisco, CA 94080 (US).			
(72) Inventors: WARD, Michael; 381 Myrtle Street, Half Moon Bay, CA 94019 (US). CLARKSON, Kathleen, A.; 53 28th Street, San Francisco, CA 94110 (US). LARENAS, Edmund, A.; 352 Caprino Way, San Carlos, CA 94070 (US). LORCH, Jeffrey, D.; 527 13th Street South, #202, Hudson, WI 54016 (US). WEISS, Geoffrey, L.; 275 Grattan Street, #6, San Francisco, CA 94117 (US).			
(74) Agent: KRUPEN, Karen, I.; Genencor International, Inc., 180 Kimball Way, South San Francisco, CA 94080 (US).			Published Without international search report and to be republished upon receipt of that report.
(54) Title: PURIFICATION AND MOLECULAR CLONING OF EG III CELLULASE			
(57) Abstract			
<p>The present invention is directed to purified EG III cellulase enzyme isolated from <i>Trichoderma longibrachiatum</i> and the amino acid sequence of the secreted (mature) and non-secreted (preprotein) forms. The present invention is further directed to the DNA fragment and sequence that encodes the EG III cellulase enzyme. Also disclosed are methods for isolating either purified or highly enriched EG III cellulase obtained from <i>Trichoderma spp.</i> or genetically modified strains of <i>Trichoderma spp.</i></p>			
BEST AVAILABLE COPY			

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	GB	United Kingdom	MR	Mauritania
AU	Australia	GE	Georgia	MW	Malawi
BB	Barbados	GN	Guinea	NE	Niger
BE	Belgium	GR	Greece	NL	Netherlands
BF	Burkina Faso	HU	Hungary	NO	Norway
BG	Bulgaria	IE	Ireland	NZ	New Zealand
BJ	Benin	IT	Italy	PL	Poland
BR	Brazil	JP	Japan	PT	Portugal
BY	Belarus	KE	Kenya	RO	Romania
CA	Canada	KG	Kyrgyzstan	RU	Russian Federation
CF	Central African Republic	KP	Democratic People's Republic of Korea	SD	Sudan
CG	Congo	KR	Republic of Korea	SE	Sweden
CH	Switzerland	KZ	Kazakhstan	SI	Slovenia
CI	Côte d'Ivoire	LI	Liechtenstein	SK	Slovakia
CM	Cameroon	LK	Sri Lanka	SN	Senegal
CN	China	LU	Luxembourg	TD	Chad
CS	Czechoslovakia	LV	Latvia	TG	Togo
CZ	Czech Republic	MC	Monaco	TJ	Tajikistan
DE	Germany	MD	Republic of Moldova	TT	Trinidad and Tobago
DK	Denmark	MG	Madagascar	UA	Ukraine
ES	Spain	ML	Mali	US	United States of America
FI	Finland	MN	Mongolia	UZ	Uzbekistan
FR	France			VN	Viet Nam
GA	Gabon				

PURIFICATION AND MOLECULAR CLONING OF EG III CELLULASE

CROSS-REFERENCE TO RELATED APPLICATIONS

This application is a continuation-in-part of U.S. Serial No. 07/862,846 filed April 3, 1992, which in turn is a continuation-in-part of U.S. Serial No. 07/707,647 filed May 30, 1991, which in turn is a continuation-in-part of U.S. Serial No. 07/668,640 filed on March 13, 1991, which in turn is a continuation-in-part of U.S. Serial No. 07/593,919 filed October 5, 1990 and now abandoned. This application is also a continuation-in-part of U.S. Serial No. 07/678,865 filed March 29, 1991 which in turn is a continuation-in-part of U.S. Serial No. 07/793,919 filed October 5, 1990 and now abandoned. The disclosure of these five applications are incorporated herein in their entirety by reference.

BACKGROUND OF THE INVENTION

1. Field of the Invention

The present invention is directed to purified EG III cellulase enzyme isolated from *Trichoderma longibrachiatum* and its amino acid sequence in both secreted and non-secreted forms. The present invention is further directed to a DNA sequence that encodes the EG III cellulase enzyme. The present invention further relates to methods of isolating purified and highly enriched EG III cellulase obtained from *Trichoderma spp.* or genetically modified strains of *Trichoderma spp.*

2. State of the Art

Cellulases are known in the art as enzymes that hydrolyze cellulose (β -1,4-glucan linkages) thereby resulting in the formation of glucose, cellobiose, cellooligosaccharides, and the like. While cellulases are produced (expressed) in fungi, bacteria and the like, cellulase produced by certain fungi, and in particular by the fungal genus *Trichoderma spp.* (especially *Trichoderma longibrachiatum*), have been given the most attention because a complete cellulase system capable of

-2-

degrading crystalline forms of cellulose is readily produced in large quantities via fermentation procedures.

In regard to the above, Wood et al, "Methods in Enzymology", 160, 25, pages 234 et seq (1988), disclose that complete fungal cellulase systems comprise several different enzyme classifications including those identified as exo-cellobiohydrolases (EC 3.2.1.91) ("CBH"), endoglucanases (EC 3.2.1.4) ("EG"), and β -glucosidases (EC 3.2.1.21) ("BG"). The fungal cellulase classifications of CBH, EG and BG can be further expanded to include multiple components within each classification. CBHs and EGs have been isolated from a variety of fungal sources.

The complete cellulase system comprising CBH, EG and BG components is required to efficiently convert crystalline cellulose to glucose. Isolated components are far less effective, if at all, in hydrolyzing crystalline cellulose. Moreover, a synergistic relationship is observed between the cellulase components, particularly if they are of different classification.

On the other hand, cellulases and components thereof, used either singularly or in combination, are also known in the art to be useful in detergent compositions, as a softening agent, and to improve the feel of cotton fabrics, and the like. However, there is a problem with using the EG I and EG II components derived from *Trichoderma* spp., and especially *Trichoderma longibrachiatum*, in detergent compositions. Specifically, such components have their maximal activity at acidic pHs whereas most laundry detergent compositions are formulated for use at neutral or alkaline (pH >7 to about 10) conditions. While it is disclosed in U.S. Serial No. 07/668,640 that the use of one or more acidic endoglucanase components of *Trichoderma longibrachiatum* in detergent compositions will provide improvements in softening, color retention/restoration and feel to cotton-containing fabrics even when treated under alkaline conditions, U.S. Serial No. 07/707,647 is directed to the discovery that the EG III component of *Trichoderma* spp. provides for superior and

-3-

unexpected advantages in detergent compositions as compared to the EG I and EG II components of *Trichoderma longibrachiatum*.

In addition to its use in laundry detergents, EG III cellulase can be used in a pre-washing step in the appropriate solution at an intermediate pH where sufficient activity exists to provide desired improvements in color retention/restoration, softening and feel as disclosed in U.S. Serial No. 07/707,647 filed May 30, 1991 and incorporated herein by reference.

EG III cellulase has a further use in the stonewashing process of colored fabrics wherein redeposition of a colorant onto the fabric may be reduced by employing purified EG III. This process is disclosed in U.S. Serial No. 07/954,113 filed September 30, 1992 and incorporated herein by reference.

Additionally, it is further contemplated that the high activity under neutral to alkaline conditions of EG III cellulase would be beneficial in textile processes for treating cotton-containing fabrics (see U.S. Serial Nos. 07/677,385 and 07/678,865 which are incorporated herein by reference in their entirety) as well as in silage and/or composting processes.

Thus, it has become of increasing interest to isolate EG III in purified form or to create a *Trichoderma* strain which secretes a cellulase product enriched for EG III for commercial use. Others in the field have described the purification of low molecular weight endoglucanases from *Trichoderma* (Shoemaker et al (1981) Trends in the Biology of Fermentations for Fuels and Chemicals (Hollaender, Rabson, Rogers, Pietro, Valentine and Wolfe, Eds.), Plenum Publishing Corp., New York; Hakansson et al, (1978) Biochim. Biophys. Acta 524:385-392; Beldman et al (1985) Eur. J Biochem. 146:301-308; Ulker and Sprey (1990) FEMS Microbiol. Lett. 69:215-220 and Sprey and Ulker (1992) FEMS Microbiol. Lett. 92:253-258). However, it is not possible to determine which, if any, represent the same protein as the EG III described herein. For example, the protein isolated by Ulker and Sprey (1990) was determined to have an arginine at its amino

-4-

terminus. However, the DNA sequence of the EG III encoding gene reported herein would predict that EG III has a glutamine residue at its amino terminus.

In light of the various applications of EG III, the present invention is directed to the complete characterization of EG III, i.e, the amino acid sequence and DNA sequence encoding EG III, purified from a fungal cellulase composition. The full characterization of EG III cellulase described herein will provide a cost-effective commercially available EG III cellulase product through genetic engineering and/or large scale protein purification procedures.

SUMMARY OF THE INVENTION

A single endoglucanase component called EG III derived from *Trichoderma spp.* has now been purified to homogeneity and the complete amino acid sequence and DNA sequence encoding this cellulase enzyme has been determined.

Accordingly, one embodiment of the present invention relates to the DNA fragment encoding EG III cellulase as shown in SEQ ID NO:11 or modification thereof characterized in having a molecular weight of about 22 to 27 Kdaltons and endoglucanase activity. A pH optimum range of about 5.5 to 6.0 has been determined for the mature and secreted form of EG III in *Trichoderma longibrachiatum*.

In another embodiment, the present invention relates to the precursor (preprotein) and mature (secreted) forms of the EG III cellulase enzyme as shown in SEQ ID NO:9 and SEQ ID NO:10, respectively, or derivatives thereof which specify the secreted EG III protein in [AS?] having a molecular weight of about 22-27 Kdaltons and endoglucanase activity.

The present invention also relates to a method for producing purified EG III cellulase enzyme from an aqueous cellulase protein mixture which can be obtained commercially or a whole cellulase composition from a wild-type *Trichoderma spp.* strain.

In another embodiment, the present invention relates to methods of isolating purified or highly enriched EG III by

-5-

using a genetically modified *Trichoderma spp.* strain wherein at least one or more exo-cellobiohydrolase components, CBH I and CBH II and endoglucanase components, EG I and EG II are inactivated for the production of EG III. More particularly, the *Trichoderma spp.* strain is genetically modified such that CBH I, CBH II, EG I and EG II are all inactivated.

A highly enriched EG III present in the protein mixture produced from the genetically modified strains of *Trichoderma spp.* described above may be acquired after subjecting the mixture to a combination of filtration and ultrafiltration steps. Alternatively, the EG III present in the protein mixture produced from the above modified strains may be further purified to homogeneity using a combination of polyethylene glycol extraction and column chromatography steps.

A further embodiment of the present invention relates to a method for producing purified or highly enriched EG III cellulase enzyme from a genetically modified *Trichoderma spp.* strain described above that in addition overexpresses EG III. EG III protein may be overproduced in a genetically modified strain described above into which multiple copies of the EG III gene have been inserted. Highly enriched EG III or purified EG III produced from an overexpressed genetically modified strain may be obtained after subjecting the protein mixture produced from this strain to concentration and/or purification procedures described above.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 shows the alignment of amino acid sequences of peptides obtained from *Trichoderma longibrachiatum* EG III with the sequences of the mature forms of endoglucanases from *Erwinia carotovora* var. *carotovora* (*E. carot.* EG) and *Aspergillus aculeatus* (*A. aculeatus* EG).

Figure 2 is a restriction map of the cloned, overlapping *Hind*III and *Asp*718 fragments of *Trichoderma longibrachiatum* genomic DNA which include the EG III encoding gene. The direction of transcription is denoted by an arrow over the EG

III coding region.

Figure 3 shows the genomic DNA sequence of the *Trichoderma longibrachiatum* EG III encoding gene. The deduced amino acid sequence of EG III is shown below the DNA sequence. An arrow denotes the deduced signal peptidase cleavage site between the last residue of the 16 amino acid signal sequence and the first residue of the mature protein. The two introns are shown in lower case letters. Sequences which match the consensus sequences thought to be involved with splicing filamentous fungal introns are underlined (Gurr et al (1987) Gene Structure in Eukaryotic Microbes (Kinghorn, Ed.), IRL Press, Oxford, UK).

DETAILED DESCRIPTION OF THE INVENTION

The present invention relates to purified EG III cellulase protein obtained from *Trichoderma* and a DNA sequence encoding the protein. In another aspect, the present invention relates to a method of isolating purified or enriched EG III cellulase obtained from *Trichoderma spp.* or genetically modified strains of *Trichoderma spp.*

Within the specification certain terms are disclosed and will be defined so as to clarify the nature of the claimed invention.

The term "EG III cellulase" refers to the endoglucanase component derived from *Trichoderma spp.* characterized by a pH optimum of about 5.5 to 6.0, an isoelectric point (pI) of from about 7.2 to 8.0 and a molecular weight of about 22 to 27 Kdaltons. Preferably, EG III cellulase is derived from either *Trichoderma longibrachiatum* or from *Trichoderma viride*. EG III cellulase derived from *Trichoderma longibrachiatum* has a pH optimum of about 5.5 to 6.0, an estimated isoelectric point (pI) of about 7.4 and an apparent molecular weight of about 22 to 27 Kdaltons as judged by polyacrylamide gel electrophoresis. EG III cellulase derived from *Trichoderma viride* has a pH optimum of about 5.5, an isoelectric point (pI) of about 7.7 and a molecular weight of about 23.5 Kdaltons.

-7-

The pH optimum of purified EG III is determined by measuring its optimal activity in a Remazol Brilliant Blue Carboxymethylcellulose (RBB-CMC) assay. The recovered EG III after each stage of purification (described in the Examples below) is determined by EG III activity using the RBB-CMC assay. EG III activity is calculated at 40°C using the following procedure.

5 to 10 μ l of recovered EG III is added at a concentration sufficient to provide the requisite amount of enzyme in the final solution. 250 μ l of 2 weight percent RBB-CMC (commercially available from MegaZyme, 6 Altona Place, North Rocks, N.S.W. 2151 Australia) is added in 0.05M citrate/phosphate buffer at a pH that may range from 4.0 to 8.0 in 0.5 pH increments. The solution is vortexed and incubated at 40°C for 30 minutes, followed by chilling in an ice bath for 5 to 10 minutes. 1000 μ l of methyl cellosolve containing 0.3M sodium acetate and 0.02M zinc acetate is added, centrifuged and the supernatant is poured into cuvettes. The optical density (OD) of the solution in each cuvette is measured at 590 nm. Higher OD levels correspond to higher levels of enzyme activity.

EG III cellulase may be purified from any strain of *Trichoderma spp.* which produces EG III under suitable fermentation conditions. While the particular source of EG III is not critical, preferred sources are *Trichoderma longibrachiatum* and *Trichoderma viride*. A particularly preferred source of EG III from *Trichoderma longibrachiatum* is Cytolase 123 cellulase which is commercially available from Genencor International, Inc., 180 Kimball Way, South San Francisco, CA 94080. Because of its high pI, EG III is found in a region of an isoelectrofocusing gel where high pI xylanases and other high pI components expressed by *Trichoderma spp.* are generally found. In fact, it has been hypothesized that the band identified as EG III was a degradation product of either EG I or II. However, gel isoelectrofocusing of EG I and EG II deleted cellulase (prepared in the manner of U.S. Serial Nos. 07/593,919 and

-8-

07/668,640) demonstrated that this band was not attributable to a degradation product of either EG I or II. (See U.S. Serial No. 07/862,846 herein incorporated by reference.)

It is noted that EG II has been previously referred to by the nomenclature "EG III" by some authors but current nomenclature uses the term "EG II". In any event, the EG II protein is substantially different from the EG III protein in its molecular weight, pI and pH optimum as evidenced by Table I of Example 2 presented below.

"Cellulase proteins" refer to cellulase proteins which contain any and all exo-cellobiohydrolase (CBH) proteins, endoglucanase (EG) proteins and β -glucosidase (BG) proteins derived from wild-type fungal sources or genetically modified microorganisms.

"Endoglucanase (EG) components" refer to the EG components of *Trichoderma spp.*, including the EG I, EG II and EG III components of *Trichoderma longibrachiatum*.

"Exo-cellobiohydrolase (CBH) components" refer to the CBH components of *Trichoderma spp.* including the CBH I and CBH II components of *Trichoderma longibrachiatum*.

Several procedures suitable for obtaining purified EG III cellulase from a complete cellulase system derived from *Trichoderma spp.* ("whole cellulase") have previously been recited in U.S. Serial Nos. 07/707,647, 07/678,865 and 07/862,846. The Examples described herein below disclose the complete purification of EG III cellulase to homogeneity by subjecting whole cellulase to purification procedures by repeated fractionation utilizing different fractionation columns preceded by an extraction step using polyethylene glycol 8000.

It is contemplated that essentially pure EG III cellulase can be prepared by genetically modifying microorganisms so as to produce enriched EG III cellulase that may subsequently be purified to homogeneity by following the purification procedure described herein.

Additionally, highly enriched EG III cellulase may be prepared by genetically modified microorganisms. The cell

-9-

culture may then be filtered to remove the cells followed by ultrafiltration to concentrate EG III. Formulation with various salts, sugars and/or preservatives may yield a commercial product. Alternatively, highly enriched EG III cellulase may be prepared from the genetically modified microorganisms described above by an extraction step using polyethylene glycol (PEG) 8000 used to further enrich EG III. The PEG may then be removed and the concentrated EG III may be formulated with salts, sugars and/or preservatives.

For example, the cellulase protein mixture for either purified EG III or the enriched EG III component may be derived from the genetically modified *Trichoderma spp.* strains wherein the genes encoding the exo-cellobiohydrolases CBH I and CBH II and endoglucanases EG I and EG II have been removed. In another example, the EG III protein could be overproduced in a strain into which multiple copies of the EG III gene have been inserted. In this case, the EG III coding region may be operably linked to a different promoter such as that from the CBH I-encoding gene. Multiple copies of the EG III encoding gene may be inserted into a strain in which the genes encoding some or all of the other secreted enzymes, e.g., cellulase or xylanase, had been inactivated.

Thus, several different sources of EG III discussed above may be employed by the methods set forth in the Examples to determine the amino acid sequence of parts or all of the EG III protein using known sequencing methods.

The present invention relates to purified EG III cellulase enzyme having a molecular weight of approximately 22-27 kD, pI of approximately 7.2 to 8.0 and pH optimum in a range of about between 5.5 to 6.0, further characterized as having the amino acid sequence shown in the appended SEQ ID NO:9 (native sequence) and SEQ ID NO:10 (proposed secreted sequence) or a derivative thereof exhibiting similar biochemical characteristics as described above and having equal to or greater than 70% sequence identity with SEQ ID NO:10. Similar biochemical characteristics of the derivative of the EG III may include pH optimum that ranges from about

5.5 to about 7.0.

The term "derivative" is intended to include derivatives of the aforementioned sequences shown by the addition of one or more amino acid residues to either or both the C- and N-terminus of the native or secreted sequence, substitution of one or more amino acid residues at one or more sites in the native or secreted sequence, deletion of one or more amino acid residues at either or both ends of the native or secreted sequence, or deletions from within or insertion of one or more amino acid residues at one or more sites in the native or secreted sequence such that a sequence identity of at least 70% with SEQ ID NO:10 is retained.

The term "preprotein or native sequence" refers to the amino acid sequence of the precursor EG III prior to cleavage of the secretory signal sequence and secretion of mature EG III outside of the cell. Thus, the preprotein amino acid sequence contains a secretion signal sequence at the N-terminus. The term "secreted or mature sequence" is the amino acid sequence of EG III minus the secretory signal sequence.

The present invention also relates to a DNA sequence from the genome of *Trichoderma longibrachiatum* that comprises a DNA sequence encoding secreted EG III cellulase as described above or a precursor form of the protein. In particular, the DNA sequence of the present invention relates to SEQ ID NO:11 in the appended Sequence Listings or a modification thereof. Examples of suitable "modifications" of the DNA sequence are nucleotide substitutions, deletions or insertions which give rise to another form of EG III having the biochemical characteristics as described above. Another example of a modified DNA sequence would be isolation of complementary DNA (cDNA) by reverse transcription of EG III mRNA by methods known in the art.

The amino acid sequence of parts of the EG III cellulase purified to homogeneity described in Example 3 below were used to design synthetic DNA probes in order to clone the gene responsible for encoding this information. The sequence of the EG III encoding gene may further be manipulated by

-11-

recognized techniques and ultimately inserted into various *Trichoderma spp.* strains or into other microorganisms to obtain higher producing organisms for commercialization. See, for example, U.S. Serial No. 07/593,919 filed October 5, 1990 and U.S. Serial No. 07/668,640 filed March 13, 1991, both of which disclose methods for genetically engineering *Trichoderma longibrachiatum* so that the modified microorganism is incapable of expressing one or more of the cellulase genes and, in fact, may overproduce another cellulase. The disclosures of both U.S. Serial No. 07/593,919 filed October 5, 1990 and U.S. Serial No. 07/668,640 filed March 13, 1991 are incorporated herein by reference in their entirety.

It has been shown that specific genes within the *Trichoderma longibrachiatum* genome can be inactivated by deletion of part or all of the gene or insertion of other DNA sequences into the gene. It is, for example, possible to create strains in which some or all of the genes encoding the major cellulases (CBH I, CBH II, EG I, or EG II) have been inactivated. As a result the culture supernatant obtained from these strains would not contain these major cellulase enzymes. In turn this would simplify purification of EG III as has previously been suggested in U.S. Serial Number 07/862,846 incorporated herein by reference.

Overproduction of EG III by *Trichoderma longibrachiatum* could be achieved by inserting multiple copies of the EG III encoding gene into the genome of this fungus (exemplified previously for EG I described in U.S. Serial No. 07/954,113 incorporated by reference in its entirety). In order to maximize production of EG III it may be desirable to operably link the EG III coding region to a highly efficient promoter region obtained from another gene such as that encoding CBH I.

Additionally, one may wish to utilize a strain of *Trichoderma longibrachiatum* in which the genes encoding the major cellulases had been inactivated as a host for overproduction of EG III. Secretion of EG III would be attained by using the coding region of EG III native preprotein having its own secretion signal. However,

-12-

secretion of mature EG III may also be possible if a different signal sequence was employed or if EG III was produced as a fusion protein attached to another secreted protein. For example, the signal sequence of CBH I could be fused with the coding region for mature EG III to allow efficient secretion of EG III. The coding region for EG III may, for example, be fused to the coding region for another cellulase (e.g., CBH I) or parts thereof, or for another secreted enzyme (e.g., protease or amylase) or parts thereof, so that a secreted fusion protein is produced.

Additionally, it would be possible to express the EG III encoding gene in other microorganisms, including, but not limited to, yeast species such as *Saccharomyces cerevisiae*, *Pichia pastoris*, *Hansenula polymorpha*, *Kluyveromyces lactis*, *Yarrowia lipolytica*, *Schanniomyces occidentalis*, etc. See, for example, PCT application Publication No. WO 85/04672. In order to obtain expression in these alternative, non-*Trichoderma* hosts, it may be necessary to functionally combine the EG III coding DNA sequence (after first removing the introns of the EG III encoding gene) with promoter and terminator sequences obtained from a gene from that particular host. It may also be necessary to substitute the DNA sequence encoding a secretion signal sequence from the alternative host for the DNA sequence encoding the EG III secretion signal sequence. Production and secretion of EG III in other organisms could enable EG III to be obtained in substantially pure form.

The cloned EG III encoding DNA could be used as a molecular probe in experiments designed to clone similar genes from other filamentous fungi. In this way it may be possible to clone genes encoding EG III-like enzymes from organisms including, but not restricted to, species of *Trichoderma*, *Humicola*, *Aspergillus*, *Neurospora*, *Acremonium* (*Chrysosporium*), *Penicillium*, *Phanaerochaete* or *Trametes*. "EG III-like" is defined herein as to describe an enzyme derived from the above genera characterized as having a pH optimum of about between 5.5 and 7.0 based on the RBB-CMC assay and having at least 50%

-13-

identity to the amino acid sequence shown in ID SEQ NO:10.

The following examples are offered to illustrate the present invention and should not be construed in any way as limiting the scope of this invention.

EXAMPLES

Example 1 demonstrates the isolation of EG III from Cytolase 123 cellulase (a complete fungal cellulase composition obtained from *Trichoderma longibrachiatum* and available from Genencor International, Inc., South San Francisco, CA) via purification procedures. A complete fungal cellulase composition containing EG III is also available commercially from other sources including those sold under the trade name of Rapidase RL™ (from Gist-brocades, Netherlands).

Additionally, complete fungal cellulase compositions may be found in fermentation cultures of *Trichoderma longibrachiatum* strains which are commercially available and on deposit at the American Type Culture Collection.

In order to enhance the efficiency of the isolation of EG III, it may be desirable to employ *Trichoderma longibrachiatum* genetically modified so as to overexpress EG III and/or to be incapable of producing one or more of EG I, EG II, CBH I and/or CBH II components. This will necessarily lead to more efficient isolation of EG III by, for example, PEG extraction as described below. Production of some of these strains of *Trichoderma longibrachiatum* are disclosed in U.S. Serial No. 07/668,640 filed March 13, 1991.

Example 1

Large Scale Extraction of EG III Cellulase Enzyme

One hundred liters of cell free cellulase filtrate were heated to about 30°C. The heated material was made about 4% wt/vol PEG 8000 (polyethylene glycol, MW of about 8000) and about 10% wt/vol anhydrous sodium sulfate. The mixture formed a two phase liquid mixture. The phases were separated using an SA-1 disk stack centrifuge. The phases were analyzed using silver staining isoelectric focusing gels. Fractionation and

-14-

enrichment were obtained for EG III and xylanase. The recovered composition contained about 20 to 50 weight percent of EG III.

Regarding the above procedure, use of a polyethylene glycol having a molecular weight substantially less than about 8000 gave inadequate separation; whereas, use of polyethylene glycol having a molecular weight substantially greater than about 8000 resulted in the exclusion of desired enzymes in the recovered composition. With regard to the amount of sodium sulfate, sodium sulfate levels substantially greater than about 10% wt/vol caused precipitation problems; whereas, sodium sulfate levels substantially less than about 10% wt/vol gave poor separation or the solution remained in a single phase.

The enriched EG III solution from the PEG extraction was diafiltered using an omega series tangential flow 8,000 ultra filtration membrane (Filtron Technology Corp., Northborough, MA) against 10 mM, pH 4.0 citrate/phosphate buffer. The solution was loaded onto an equilibrated (pH 4.0, 10 mM citrate/phosphate) SP Trisacryl column. The EG III component was eluted with 250 mM sodium chloride.

Example 2

Purification of EG III to Homogeneity Via Fractionation

In order to isolate EG III to a level of homogeneity to perform amino acid sequence analysis, the EG III composition described in Example 1 was further subjected to column chromatography. The further fractionation was done on a FPLC system using a Mono-S-HR 5/5 column (available from Pharmacia LKB Biotechnology, Piscataway, NJ). The FPLC system consists of a liquid chromatography controller, 2 pumps, a dual path monitor, a fraction collector and a chart recorder (all of which are available from Pharmacia LKB Biotechnology, Piscataway, NJ). The fractionation was conducted by desalting 5 ml of the EG III sample prepared in Example 1 with a 20 ml Sephadex G-25 column which had been previously equilibrated with 10 mM sodium citrate pH 4. The solution was loaded onto

-15-

mono-S-HR 5/5 column previously equilibrated with 10 mM sodium citrate pH 4.0 and eluted with 0-200 mM aqueous gradient of NaCl at 1%/min with a flow rate of 0.5 ml/minute with samples collected in 1 ml fractions. EG III was recovered in two fractions and was determined to be greater than 90% pure by SDS gel electrophoresis.

EG III purified in the above manner has the following characteristics which are compared to the other endoglucanases isolated from *Trichoderma longibrachiatum*.

TABLE I

	MW	pI ¹	pH optimum ²
EG I	~47-49 kD	4.7	~5
EG II	~35 kD	5.5	~5
EG III	~22-27 kD	7.4	~5.5-6.0

1. pI value is an estimate based on isoelectrofocusing gels.
2. pH optimum determined by RBB-CMC activity.

As can be seen from the above table, EG III has both a higher pH optimum and a higher pI as compared to the other endoglucanase components of *Trichoderma longibrachiatum*. It has also been shown that EG III retains significant RBB-CMC activity under alkaline pHs (disclosed in U.S. Serial No. 07/862,846).

Likewise, EG III cellulase from other strains of *Trichoderma spp.* can be purified in the same manner as described above. For example, EG III cellulase derived from *Trichoderma viride* has been described by Voragen et al, Methods in Enzymology, 160:243-249. This reference describes the EG III cellulase as having a molecular weight of about 23.5 Kdaltons, a pH optimum of 5.5, and a pI of 7.7.

Example 3Amino Acid Sequence Determination of EG III

Purified EG III was cleaved to produce smaller peptides by treatment with either cyanogen bromide or trypsin as follows. First, the EG III was precipitated by adding 900 microliters of acetone to 100 microliters of 1 mg/ml solution of EG III. After incubation at -20°C for 10 minutes the precipitated EG III was collected by centrifugation and the pellet was dried. For cyanogen bromide treatment the EG III pellet was dissolved in 100 microliters of 6M urea in 88% formic acid; 10 microliters of 200 mg/ml solution of cyanogen bromide was added and the mixture incubated for 4 hours at 25°C. For trypsin treatment the EG III pellet was dissolved in 50 microliters of Tris (pH 8.0), 2M urea, 0.5% trifluoroacetic acid (TFA); 5 micrograms of trypsin was added and the mixture incubated at 37°C for 4 hours.

The resulting peptides were individually purified by high pressure liquid chromatography (HPLC) as follows. A Synchropak RP-4 column was equilibrated in deionized water with 0.5% triethylamine (TEA) and 0.5% trifluoroacetic acid. The sample was loaded and then eluted with a gradient of 99% acetonitrile, 0.5% TEA, 0.5% TFA at 1% per minute.

The amino acid sequences of the amino-terminal regions of the peptides were determined by the method of Edman using a fully automated apparatus (Edman, P. and Begg, G. (1967) Eur. J. Biochem. 1:80-91). The sequences obtained are shown below. Since cyanogen bromide is known to cleave proteins after methionine residues both peptide 1 and peptide 2 shown below would be expected to be preceded by methionine in the intact protein.

Peptide 1 (ID SEQ NO:1) obtained by cyanogen bromide cleavage:

IWLGLKYVDGPIGSSQGTNVNKGQXXL

Peptide 2 (ID SEQ NO:2) obtained by cyanogen bromide cleavage:

-17-

PTTASWSYSGSNIRANVAYDLFTAAN

Peptide 3 (ID SEQ NO:3) obtained by trypsin cleavage:
TVNSISSMPTTASW

Peptide 4 (ID SEQ NO:4) obtained by trypsin cleavage:
NFFNYLR

Peptide 5 (ID SEQ NO:5) obtained by trypsin cleavage:
SYQNSQIAI

The peptide sequences shown above were compared to the known amino acid sequence of endoglucanases from *Aspergillus aculeatus* (Ooi et al (1990) Curr. Genet. 18:217-222; Ooi et al (1990) Nucl. Acids Res. 18:5884) and *Erwinia carotovora* subsp. *carotovora* (Saarilahti et al, Gene 90:9-14) and similarities were observed (Figure 1).

Example 4

Cloning the Gene Encoding EG III

Three degenerate pools of oligonucleotides were synthesized according to the sequences given below. One of these pools (1) was designed to include an *EcoRI* restriction site at the 5' end and to contain all possible DNA sequences which could encode the amino acid sequence ANVAYD, with only the first two nucleotides of the aspartate codon being used. The other two pools were designed to have a *PstI* restriction site at the 5' end and between them to contain the reverse complement of all possible DNA sequences encoding the amino acid sequence ELMIWL, with only the first two nucleotides of the terminal leucine codon being used.

Pool 1 (256 different 27mers) (ID SEQ NO:6)
CGCGGAATTTCGC(N)AA(C/T)GT(N)GC(N)TA(C/T)GA

Pool 2 (48 different 25mers) (ID SEQ NO:7)
ATCTGCAGA(A/G)CCA(A/G/T)ATCAT(N)AG(T/C)TC

-18-

Pool 3 (24 different 25mers) (ID SEQ NO:8)
ATCTGCAGA(A/G)CCA(A/G/T)ATCAT(C/T)AA(T/C)TC

It was postulated that these primers could be used in pairs (either 1 with 2a, or 1 with 2b) in order to amplify an approximately 100 bp *T. longibrachiatum* DNA fragment using a polymerase chain reaction (PCR) procedure.

Complementary DNA (cDNA) was prepared using a commercial kit (Invitrogen Corp., San Diego, CA) from poly(A) RNA extracted from a culture of *T. longibrachiatum* induced for cellulase production by growth with crystalline cellulose as the sole carbon source. PCR was performed with this cDNA as a template and using the oligonucleotides described above as primers. The best combination of oligonucleotides appeared to be Pool 1 used in combination with Pool 2a in that the highest abundance of a 100 bp DNA fragment was amplified using these as primers. This 100 bp DNA fragment was gel purified, digested with *EcoRI* and *PstI*, and ligated with M13mp19 DNA which had also been cut with *EcoRI* and *PstI*. Following transfection into *E. coli* single-stranded DNA was isolated from individual M13 plaques for DNA sequence analysis. Clones were identified which contained a 100 bp DNA fragment, the sequence of which suggested that it encoded the expected portion of EG III protein.

Total DNA was extracted from *T. longibrachiatum* strain RL-P37 and digested with various restriction enzymes. The digested DNA was subjected to agarose gel electrophoresis, transferred to a nylon membrane and hybridized with radio-labelled DNA from an M13 clone containing the 100 bp PCR fragment. From this Southern analysis it was determined that the EG III coding region resided on a 3 kb *Asp718* fragment of genomic DNA. A gene library was constructed by size fractionating (around 3 kb) *Asp718* digested RL-P37 genomic DNA and ligating this with *Asp718* digested pUC219. Plasmid pUC219 is derived from pUC119 (described by Wilson et al (1989) Gene 77:69-78) by expanding the multiple cloning site to include

restriction sites for *Bgl*III, *Cla*I and *Xho*I.

Following transformation of the ligated DNA into *E. coli* cells and selection for ampicillin resistance the resulting colonies were screened by colony hybridization using the 100 bp PCR fragment as a probe. DNA was isolated from the positive colonies and a restriction map was generated (see Figure 2). It was noted that the 100 bp PCR fragment of EG III hybridized very near to one end of the cloned 3 kb *Asp*718 fragment containing the EG III encoding gene. As a result it was necessary to clone an overlapping *Hind*III fragment of RL-P37 genomic DNA in order to obtain the entire gene for EG III.

The DNA sequence of the EG III coding region and some of the flanking DNA was determined and is shown in Figure 3. Based on similarity between this DNA sequence and that of the *A. aculeatus* endoglucanase gene disclosed by Ooi et al, Applicants were able to distinguish the start and end of the coding region and to identify two introns which interrupt the coding region. These introns contain sequences at their 5' and 3' ends as well as an internal 6 bp sequence which match the consensus sequences considered to be important for splicing fungal introns. Applicants deduce that EG III is initially synthesized as a preprotein having a 16 amino acid secretion signal sequence at the amino-terminus. If this is the case, the final three amino acids of the signal sequence would be Ala Leu Ala, which is in agreement with the consensus cleavage site for signal peptidase which serves to remove the signal sequence during protein secretion, see Perlman, D. and Halvorson, H.O. (1983) *J. Mol. Biol.* 167: 391-409. Cleavage by signal peptidase to remove the signal sequence would yield a protein of 219 amino acids with the amino-terminal residue being Gln. In contrast, the low molecular weight endoglucanase purified by Ulker and Sprey ((1990) *FEMS Microbiol. Lett.* 69:215-220) was thought to have Arg as the amino terminal residue and to have approximately 235 amino acids. The endoglucanase purified by Hakansson et al was judged to contain 197 amino acid residues.

* * * * *

-20-

While the foregoing invention has been described in some detail for the purposes of clarity and understanding, it will be appreciated by one skilled in the art from a reading of this disclosure that various changes in form and detail can be made without departing from the true scope of the invention and appended claims.

-21-

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT: Ward, Michael
Clarkson, Kathleen A.
Weiss, Geoffrey L.
Larenas, Edmund
Lorch, Jeffrey D.

(ii) TITLE OF INVENTION: Purification and Molecular Cloning of EG
III Cellulase

(iii) NUMBER OF SEQUENCES: 11

(iv) CORRESPONDENCE ADDRESS:

(A) ADDRESSEE: Genencor International
(B) STREET: 180 Kimball Way
(C) CITY: South San Francisco
(D) STATE: CA
(E) COUNTRY: USA
(F) ZIP: 94080

(v) COMPUTER READABLE FORM:

(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.25

(vi) CURRENT APPLICATION DATA:

(A) APPLICATION NUMBER:
(B) FILING DATE:
(C) CLASSIFICATION:

(viii) ATTORNEY/AGENT INFORMATION:

(A) NAME: Krupen, Karen
(B) REGISTRATION NUMBER: 34,647

-22-

(ix) TELECOMMUNICATION INFORMATION:

- (A) TELEPHONE: 415 742-7218
- (B) TELEFAX: 415 742-7217

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 28 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

Ile Trp Leu Gly Lys Tyr Gly Asp Ile Gly Pro Ile Gly Ser Ser Gln
1 5 10 15

Gly Thr Val Asn Val Gly Gly Gln Xaa Xaa Xaa Leu
 20 25

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 26 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

-23-

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

Pro Thr Thr Ala Ser Trp Ser Tyr Ser Gly Ser Asn Ile Arg Ala Asn
1 5 10 15

Val Ala Tyr Asp Leu Phe Thr Ala Ala Asn
 20 25

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 14 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

Thr Val Asn Ser Ile Ser Ser Met Pro Thr Thr Ala Ser Trp
1 5 10

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 7 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

-24-

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

Asn Phe Phe Asn Tyr Leu Arg
1 5

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 9 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

Ser Tyr Gln Asn Ser Gln Ile Ala Ile
1 5

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 27 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

CGCGGAATTC GCNAAYGTNG CNTAYGA

27

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 25 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

ATCTGCAGAR CCADATCATN AGYTC

25

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 25 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

ATCTGCAGAR CCADATCATY AAYTC

25

-26-

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 234 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

```

Met Lys Phe Leu Gln Val Leu Pro Ala Leu Ile Pro Ala Ala Leu Ala
1           5           10           15

Gln Thr Ser Cys Asp Gln Trp Ala Thr Phe Thr Gly Asn Gly Tyr Thr
          20           25           30

Val Ser Asn Asn Leu Trp Gly Ala Ser Ala Gly Ser Gly Phe Gly Cys
          35           40           45

Val Thr Ala Val Ser Leu Ser Gly Gly Ala Ser Trp His Ala Asp Trp
          50           55           60

Gln Trp Ser Gly Gly Gln Asn Asn Val Lys Ser Tyr Gln Asn Ser Gln
65           70           75           80

Ile Ala Ile Pro Gln Lys Arg Thr Val Asn Ser Ile Ser Ser Met Pro
          85           90           95

Thr Thr Ala Ser Trp Ser Tyr Ser Gly Ser Asn Ile Arg Ala Asn Val
          100          105          110

```

-27-

Ala Tyr Asp Leu Phe Thr Ala Ala Asn Pro Asn His Val Thr Tyr Ser
115 120 125

Gly Asp Tyr Glu Leu Met Ile Trp Leu Gly Lys Tyr Gly Asp Ile Gly
130 135 140

Pro Ile Gly Ser Ser Gln Gly Thr Val Asn Val Gly Gly Gln Ser Trp
145 150 155 160

Thr Leu Tyr Tyr Gly Tyr Asn Gly Ala Met Gln Val Tyr Ser Phe Val
165 170 175

Ala Gln Thr Asn Thr Thr Asn Tyr Ser Gly Asp Val Lys Asn Phe Phe
180 185 190

Asn Tyr Leu Arg Asp Asn Lys Gly Tyr Asn Ala Ala Gly Gln Tyr Val
195 200 205

Leu Ser Tyr Gln Phe Gly Thr Glu Pro Phe Thr Gly Ser Gly Thr Leu
210 215 220

Asn Val Ala Ser Trp Thr Ala Ser Ile Asn
225 230

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 218 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

-28-

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

Gln	Thr	Ser	Cys	Asp	Gln	Trp	Ala	Thr	Phe	Thr	Gly	Asn	Gly	Tyr	Thr
1				5					10					15	

Val	Ser	Asn	Asn	Leu	Trp	Gly	Ala	Ser	Ala	Gly	Ser	Gly	Phe	Gly	Cys
			20					25					30		

Val	Thr	Ala	Val	Ser	Leu	Ser	Gly	Gly	Ala	Ser	Trp	His	Ala	Asp	Trp
		35					40					45			

Gln	Trp	Ser	Gly	Gly	Gln	Asn	Asn	Val	Lys	Ser	Tyr	Gln	Asn	Ser	Gln
	50					55					60				

Ile	Ala	Ile	Pro	Gln	Lys	Arg	Thr	Val	Asn	Ser	Ile	Ser	Ser	Met	Pro
65					70					75				80	

Thr	Thr	Ala	Ser	Trp	Ser	Tyr	Ser	Gly	Ser	Asn	Ile	Arg	Ala	Asn	Val
				85					90					95	

Ala	Tyr	Asp	Leu	Phe	Thr	Ala	Ala	Asn	Pro	Asn	His	Val	Thr	Tyr	Ser
			100					105					110		

Gly	Asp	Tyr	Glu	Leu	Met	Ile	Trp	Leu	Gly	Lys	Tyr	Gly	Asp	Ile	Gly
		115					120					125			

Pro	Ile	Gly	Ser	Ser	Gln	Gly	Thr	Val	Asn	Val	Gly	Gly	Gln	Ser	Trp
		130				135					140				

Thr	Leu	Tyr	Tyr	Gly	Tyr	Asn	Gly	Ala	Met	Gln	Val	Tyr	Ser	Phe	Val
145					150					155				160	

Ala	Gln	Thr	Asn	Thr	Thr	Asn	Tyr	Ser	Gly	Asp	Val	Lys	Asn	Phe	Phe
			165						170					175	

SUBSTITUTE SHEET (RULE 26)

-29-

Asn Tyr Leu Arg Asp Asn Lys Gly Tyr Asn Ala Ala Gly Gln Tyr Val
 180 185 190

Leu Ser Tyr Gln Phe Gly Thr Glu Pro Phe Thr Gly Ser Gly Thr Leu
 195 200 205

Asn Val Ala Ser Trp Thr Ala Ser Ile Asn
 210 215

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1050 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

GGGTGGTCTG GATGAAACGT CTTGGCCAAA TCGTGATCGA TTGATACTCG CATCTATAAG	60
ATGGCACAGA TCGACTCTTG ATTCACAGAC ATCCGTCAGC CCTCAAGCCG TTGCAAGTC	120
CACAAACACA AGCACAAGCA TAGCGTCGCA ATGAAGTTCC TTCAAGTCCT CCCTGCCCTC	180
ATACCGGCCG CCCTGGCCCA AACCAGCTGT GACCAAGTGG CAACCTTCAC TGGCAACGGC	240
TACACAGTCA GCAACAACCT TTGGGGAGCA TCAGCCGGCT CTGGATTGTT CTGCGTGACG	300
GCGGTATCGC TCAGCGGCGG GGCCTCCTGG CACGCAGACT GGCAGTGGTC CGGCGGCCAG	360
AACAACGTCA AGTCGTACCA GAACTCTCAG ATTGCCATTC CCCAGAAGAG GACCGTCAAC	420
AGCATCAGCA GCATGCCCAC CACTGCCAGC TGGAGCTACA GCGGGAGCAA CATCCGCGCT	480
AATGTTGCGT ATGACTTGTT CACCGCAGCC AACCCGAATC ATGTCACGTA CTCGGGAGAC	540
TACGAACTCA TGATCTGGTA AGCCATAAGA AGTGACCCTC CTTGATAGTT TCGACTAACA	600

SUBSTITUTE SHEET (RULE 26)

-30-

ACATGTCTTG AGGCTTGGCA AATACGGCGA TATTGGGCCG ATTGGGTCCT CACAGGGAAC	660
AGTCAACGTC GGTGGCCAGA GCTGGACGCT CTACTATGGC TACAACGGAG CCATGCAAGT	720
CTATTCCTTT GTGGCCCAGA CCAACACTAC CAACTACAGC GGAGATGTCA AGAACTTCTT	780
CAATTATCTC CGAGACAATA AAGGATACAA CGCTGCAGGC CAATATGTTC TTAGTAAGTC	840
ACCCTCACTG TGA CTGGGCT GAGTTTGTTG CAACGTTTGC TAACAAAACC TTCGTATAGG	900
CTACCAATTT GGTACCGAGC CCTTCACGGG CAGTGGA ACT CTGAACGTCG CATCCTGGAC	960
CGCATCTATC AACTAAAACC TGGAACGTG AGATGTGGTG GGCATACGTT ATTGAGCGAG	1020
GGAAAAAAG CATTGGATCC ATTGAAGATG	1050

SUBSTITUTE SHEET (RULE 26)

-31-

WHAT IS CLAIMED IS

1. A DNA sequence encoding EG III enzyme comprising the DNA sequence shown in ID SEQ NO:11, or a modification thereof.
2. An EG III enzyme which enzyme has the amino acid sequence shown in ID SEQ NO:9 or a derivative thereof.
3. An EG III enzyme exhibiting endoglucanase activity which enzyme has the amino acid sequence shown in ID SEQ NO:10, or a derivative thereof exhibiting endoglucanase activity as determined by a RBB-CMC assay.
4. An EG III-like enzyme having the amino acid sequence containing at least 50% homology to the sequence of ID SEQ NO:10 and exhibiting endoglucanase activity as determined by a RBB-CMC assay and having a pH optimum of about 5.5 to 7.0.
5. An EG III enzyme exhibiting endoglucanase activity which enzyme has the amino acid sequence as shown in ID SEQ NO:10 or a derivative thereof and an additional signal sequence.
6. A fusion protein, said protein having the amino acid sequence as shown in ID SEQ NO:10 or derivative thereof and a secretory protein or part thereof fused to said sequence.
7. A method for isolating purified EG III cellulase having the sequence as shown in ID SEQ NO:10 comprising the steps of:
 - a) adding an effective amount of polyethylene glycol to an aqueous mixture of cellulase proteins under conditions wherein substantially all of the cellulase proteins, other than EG III, are retained in an EG III-poor aqueous phase and the EG III cellulase component is retained in an EG III-rich polyethylene glycol phase;
 - b) removing the polyethylene glycol from the EG

-32-

III-rich polyethylene glycol phase; and

c) subjecting filtered EG III to column chromatography to obtain purified EG III.

8. The method of claim 6 wherein said aqueous mixture is a cell-free cellulase mixture.

9. The method of claim 6 wherein said aqueous mixture is obtained from the fermentation culture of *Trichoderma spp.* strain that has been genetically modified such that at least one or more exo-cellobiohydrolase or endoglucanase genes have been inactivated.

10. The method of claim 6 wherein said aqueous mixture is obtained from the fermentation culture of *Trichoderma spp.* strain that has been genetically modified such that CBH I, CBH II, EG I and EG II have been inactivated.

11. The method of claim 10 wherein said *Trichoderma spp.* strain has additional EG III genes inserted in the genome.

12. The method of claim 7 wherein an inorganic salt is added to the aqueous mixture before the addition of the polyethylene glycol.

13. A method for obtaining highly enriched EG III cellulase having the sequence as shown in ID SEQ NO:10 or derivative thereof comprising the steps of:

a) preparing a fermentation culture derived from a genetically modified *Trichoderma* species strain wherein the genes encoding at least one or more CBH I, CBH II, EG I and EG II have been inactivated;

b) filtering said fermentation culture; and

c) ultra-filtrating the fermentation culture after step b).

14. A method for obtaining highly enriched EG III

-33-

cellulase having the sequence as shown in ID SEQ NO:10 or derivative thereof comprising the steps of:

- a) preparing a fermentation culture derived from a genetically modified *Trichoderma* species strain wherein the genes encoding CBH I, CBH II, EG I and EG II have been inactivated;
- b) filtering said fermentation culture; and
- c) ultra-filtrating the fermentation culture after step b).

15. The method of claim 13 or 14 wherein said genetically modified *Trichoderma* species strain has been additionally modified to overexpress EG III.

16. A method for obtaining highly enriched EG III cellulase having the sequence as shown in ID SEQ NO:10 or derivative thereof comprising the steps of:

- a) preparing a fermentation culture derived from a genetically modified *Trichoderma* species strain wherein the genes encoding at least one or more CBH I, CBH II, EG I and EG II have been inactivated;
- b) adding an effective amount of polyethylene glycol to an aqueous mixture of cellulase proteins under conditions wherein substantially all of the cellulase proteins, other than EG III, are retained in an EG III-poor aqueous phase and the EG III cellulase component is retained in an EG III-rich polyethylene glycol phase; and
- c) removing the polyethylene glycol from the EG III-rich polyethylene glycol phase.

17. A method for obtaining highly enriched EG III cellulase having the sequence as shown in ID SEQ NO:10 or derivative thereof comprising the steps of:

- a) preparing a fermentation culture derived from a genetically modified *Trichoderma* species strain wherein the genes encoding CBH I, CBH II, EG I and EG II have been inactivated;

-34-

b) adding an effective amount of polyethylene glycol to an aqueous mixture of cellulase proteins under conditions wherein substantially all of the cellulase proteins, other than EG III, are retained in an EG III-poor aqueous phase and the EG III cellulase component is retained in an EG III-rich polyethylene glycol phase; and

c) removing the polyethylene glycol from the EG III-rich polyethylene glycol phase.

18. The method of claim 16 or 17 wherein said genetically modified *Trichoderma* species strain has been additionally modified to overexpress EG III.

19. The method of claim 16 or 17 wherein an inorganic salt is added to the aqueous mixture before the addition of the polyethylene glycol.

E. carot. EG
A. aculeatus EG
ASSNDADKLYFGNNKYLFNNVWGDEIKGQ-QTIFYNSPISMGW--NWHWPSSTHSVKAYPSLVSGW
QQAQLCDQYATYTGGVYTTNNNLNGKDAGSGSQCTTVNSASSAGTSWSTKWNWGGGENSVKSYAN-----
SYQN-----

E. carot. EG
A. aculeatus EG
HWTAGYTENSGLPQLSSNKSITSNVYTSIKATGTYNAAVDIWEHTDKANWDSSPTDELMWLND-TNA
---SGLTFNKKLVSQIS-QIPTTARWSYDNTGIRA-DVAYDL-FTAADINHVTWSGDYELMIWLARYGGV
---SQIAI MPTTASWSYSGSNIRA-NVAYDL-FTAAN MIWLKYGDI
TVNSISS-MPTTASW

E. carot. EG
A. aculeatus EG
GPAGDYIETVFLGDSNNVFKGWINADNGGWNVFSFVHTSGTNSASLNIRHFTDYLVTQKQWMSDEKYI
QPIGSOIATATVDGOTWEL---WYGA--NGSQKTYSEVAPTPTSFGDVNDEFKYLTONHGFEPASSQYL
GPIGSSQGTNVNNGGQXXL NFFNYLR

E. carot. EG
A. aculeatus EG
SSVEFGTEIF-GGDGQIDITEWRVDVK
ITLQFGTEPFTGGPATLSVSNWSASVQ

Figure 1



Figure 2

GGGTGGTCTGGATGAAACGCTCTTGGCCAAATCGTGATCGATTGATACTCGCATCTATAAGATGGCACAGA 70
TCGACTCTTGATTCACAGACATCCGTAGCCCTCAAGCCGTTTGCAAGTCCACAAACACAAGCACAAAGCA 140
TAGCGTCGCAATGAAGTTCCTTCAAGTCCCTGCCCTCATACGGCCGCCCTGGCCCAAACCAGCTGT 210
MetLysPheLeuGlnValLeuProAlaLeuIleProAlaAlaLeuAlaGlnThrSerCys
↑
GACCAGTGGGCAACCTTCACTGGCAACGGCTACACAGTCAGCAACAACCTTGGGGAGCATCAGCCGGCT 280
AspGlnTrpAlaThrPheThrGlyAsnGlyTyrThrValSerAsnAsnLeuTrpGlyAlaSerAlaGly
CTGGATTGGCTGCGTGACGGCGGTATCGCTCAGCGGCGGGCCCTCCTGGCACGCAGACTGGCAGTGGTC 350
SerGlyPheGlyCysValThrAlaValSerLeuSerGlyGlyAlaSerTrpHisAlaAspTrpGlnTrpSer
CGGCGGCCAGAACAACGTCAGTACGACAGAACTCTCAGATTGCCATTCCCCAGAAGAGGACCGTCAAC 420
GlyGlyGlnAsnAsnValLysSerTyrGlnAsnSerGlnIleAlaIleProGlnLysArgThrValAsn
AGCATCAGCAGCATGCCCACCACTGCCAGCTGGAGCTACAGCGGGAGCAACATCCGCGCTAATGTTGCGT 490
SerIleSerSerMetProThrThrAlaSerTrpSerTyrSerGlySerAsnIleArgAlaAsnValAla
ATGACTTGTTTACCAGCAACCCGAATCATGTACGTACTCGGGAGACTACGAACTCATGATCTGgta 560
TyrAspLeuPheThrAlaAlaAsnProAsnHisValThrTyrSerGlyAspTyrGluLeuMetIleTrp
agccataagaagtgaccctccttgatagtttcgactaacaacatgtcttgagGCTTGGCAAATACGGCGA 630
LeuGlyLysTyrGlyAsp
TATTGGGCCGATTGGGTCCTCACAGGGAACAGTCAACGTCGGTGGCCAGAGCTGGACGCTCTACTATGGC 700
IleGlyProIleGlySerSerGlnGlyThrValAsnValGlyGlyGlnSerTrpThrLeuTyrTyrGly
TACAACGGAGCCATGCAAGTCTATTCTTTGTGGCCAGACCAACACTACCAACTACAGCGGAGATGTCA 770
TyrAsnGlyAlaMetGlnValTyrSerPheValAlaGlnThrAsnThrThrAsnTyrSerGlyAspVal
AGAACTTCTTCAATTATCTCCGAGACAATAAAGGATACAACGCTGCAGGCCAATATGTTCTTAgtaaagtc 840
LysAsnPhePheAsnTyrLeuArgAspAsnLysGlyTyrAsnAlaAlaGlyGlnTyrValLeu
accctcactgtgactgggctgagtttggttgcaacgtttgctaacaacaccttcgtatagGCTACCAATTT 910
SerTyrGlnPhe
GGTACCGAGCCCTTACGGGCAGTGGAACTCTGAACGTCGCATCCTGGACCGCATCTATCAACTAAACC 980
GlyThrGluProPheThrGlySerGlyThrLeuAsnValAlaSerTrpThrAlaSerIleAsn***
TGGAAACGTGAGATGTGGTGGGCATACGTTATTGAGCGAGGGAAAAAAGCATTGGATCCATTGAAGATG 1050

Figure 3
SUBSTITUTE SHEET (RULE 26)